DOCUMENT RESUME

ED 291 754 TM 011 039

AUTHOR Stanley, Julian C.
TITLE Study of Mathematically Precocious Youth (SMPY). Summary of Points Made in the Symposium.
PUB DATE Apr 87
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Reports - General (140) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Advanced Placement; Aptitude Tests; College Entrance Examinations; *Mathematics Achievement; *Sex Differences; *Standardized Tests; *Test Bias
IDENTIFIERS Graduate Record Examinations; *Study of Values

ABSTRACT

This paper is an overview of some points made at the Annual Meeting of the American Educational Research Association in April of 1987. Gender effects were computed on 82 nationally standardized tests designed to determine precocity among youth. The effect sizes ranged from a magnitude of 0.50 (favoring females) for spelling in grade 12 on the Differential Aptitude Tests (DATs) to 0.89 (favoring males) for mechanical reasoning on the DATs in grade 12. The largest effect size on any of the other 80 tests was 0.76 (favoring males) for the advanced examination in political science of the Graduate Record Examinations. The results of this research indicate that there was a strong tendency for tests taken mainly by males to yield the largest effect sizes favoring males and for tests taken mainly by females to yield small effect sizes, some of which favored females. All of the tests examined, except the DATs, are used primarily for selection or awarding of advanced standing in college. Although research indicates that girls and young women tend to be better students than do boys and young men, female students tend to be outperformed by male students on most standardized tests. Study results also indicate that women seem more oriented toward social, aesthetic, and religious subject matter, while men seem more interested in science, practicality, conspicuous consumption, power, and control. The Allport-Vernon-Lindzey inventory of evaluative attitudes might help researchers understand females' preferences and subject-matter orientations. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Julian Stanley
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Summary of Points Made in the Symposium*

Julian C. Stanley

Study of Mathematically Precocious Youth (SMPY)

Johns Hopkins University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED291754

Brody, Dauber, Lupkowski, and Stanley have computed gender effect sizes for 82 nationally standardized tests, some of them over several years or grades. Benbow has studied gender differences among highly able youths. Her results are important, but for uniformity and to save time for the discussants they are not covered in the present summary. Dr. Benbow's pioneering articles in Science, the American Educational Research Journal, and other professional journals provided a main post-Maccoby impetus for the study of gender differences in performance on cognitive tests.

Most of the comparisons involve large numbers of examinees, so questions of statistical significance are largely irrelevant. The effect sizes ranged from a magnitude of .50, favoring females, for Spelling in the twelfth grade on the Differential Aptitude Tests, to .89, favoring males, for Mechanical Reasoning on that same battery in the same grade. The largest effect size on any of the other 80 tests was .76, favoring males, for the advanced examination in Political Science of the Graduate Record Examinations. The second largest was .71, also favoring males, for the GRE advanced examination in Mathematics. Aside from three Differential Aptitude Tests, the largest effect size favoring females was .19 on the Verbal part of the Graduate

TM 011 039

*Paper read at the annual meeting of the American Educational Research Association in Washington, D.C., on Friday, 24 April 1987, during the 8:15-10:15 a.m. symposium entitled "Sex Differences in Cognitive Abilities and Achievement" in the Diplomat Room, West Floor of the Omni Shoreham Hotel.

Management Admissions Tests.

There was a strong tendency for tests taken mainly by males to yield the largest effect sizes favoring males, and for tests taken mainly by females to yield small effect sizes, some of which favored females. Females scored better than males only on certain verbal measures and Clerical Speed and Accuracy, but, except for the DAT, not as high as .20 on any test. Males scored considerably higher than females on all six physics tests (.59, .56, .51, .41, .37, .29), and all four chemistry tests (.57, .39, .37, .33). They scored higher on 12 of the 16 quantitative tests, (.71, .67, .43, .41, .39, .38, .37, .34, .20, .18, .04, .04); females scored higher (magnitude .16, .16, .15, .01) on four of the five school years of the DAT Numerical Ability test. Males scored considerably higher on computer science (.62, .50), biology (.36, .29, .25, .22), science (.40, .38), history and political science (.76, .58, .57, .40, .26, .24), and music (.56, .27, .05).

Let's examine the collective results for history and political science [Show the transparency.] All but two of the six effect sizes are of considerable magnitude. The median effect size exceeds that for physics and for quantitative ability, but is less than for computer science. Most surprising to me in this study were the results for history and computer science. Somehow, I had expected females not to score much, if any, below males on standardized history tests. Also, my associates and I had not anticipated the large sex differences in computer science, especially not on the Advanced Placement examination.

Clearly, across this wide domain of types and levels of tests males tend to score appreciably, and sometimes much, higher than females. The most differences favoring females occurred in Grades 8 through 12 on the DAT, which was standardized on a presumably representative national sample of students in

those grades. In what may be the most highly selected group in this study, those persons taking the 17 advanced subject-matter tests of the Graduate Record Examinations, differences uniformly and often strikingly favored males (range .16 to .76, median .48). The GRE Quantitative aptitude effect size was .67, also favoring males. As Brody showed, the females had an effect-size advantage on only one of the 30 graduate admissions tests she studied: .19 on the verbal part of the Graduate Management Admissions Test. Note carefully, however, that there was little difference on the Law School Admissions Test any of the five years observed.

The higher scoring of females on the DAT held up well over the five grades (8-12) on Spelling, Language Usage, and Clerical Speed and Accuracy, but declined on Numerical Ability, Abstract Reasoning, and Verbal Reasoning. Differences favoring males rose grade by grade on Mechanical Reasoning, from .66 to .89, and to a lesser extent on Space Relations (.13, .08, .12, .20, .22).

All of the tests examined except the DAT are used primarily for selection or awarding of advanced standing in college. This means that most colleges and universities receiving these test scores are little concerned with the degree of overlap of the score distributions of males vs. females. Instead, they are seeking scores high enough to meet their entrance or placement criteria. Thus, although the effect-size statistic provides uniformity of comparison of various tests in this symposium, it does not address the advantage or disadvantage that one sex may have over the other in a particular situation. Usually, the magnitude of the effect size is probably related fairly closely to selection or placement advantage, but glaring exceptions occur. For example, on the DAT Spelling test females systematically excelled males with respect to effect size (magnitude .39, .38, .48, .47, and .50), but little at the 99th percentile of the test's national

norms. Several of the speakers have illustrated this difference by using gender ratios at selected cutting points such as more than 3 on AP examinations. What Cohen calls small or even negligible effect sizes may produce sizable odds in the upper part of the distribution. It seems likely to me that the two discussants will go further with this contrast between effect sizes and ratios.

All in all, it is a perplexing picture. We know from much other research that girls and young women tend to be better students in school than boys and young men are. Why should they be outperformed by males on most standardized tests, especially those elected to be taken by few females? Would it serve any useful purpose to assert, vaguely and perhaps defiantly, that all but verbally oriented standardized tests are "biased" against females? The more appropriate question would seem to be: "Why did the differences occur?" We of SMPY do not know why. We have tried here to firm up the "whats" of the situation so that other researchers may pursue the "whys." Only by resolutely examining the discrepancies exposed here and elsewhere can one hope to find ameliorative procedures for helping young women prepare well for careers in science, mathematics, and (it would appear) history and political science.

The Allport-Vernon-Lindzey "A Study of Values" (SV) inventory of evaluative attitudes might help researchers understand females' preferences and subject-matter orientations. Women tend to score appreciably higher than men on the social, aesthetic, and religious scales of the SV and appreciably lower than men on the theoretical, economic, and political scales. For whatever reasons, women seem more oriented toward social service, social interaction, and aesthetic and religious considerations than men are. Men seem more interested in science, practicality, conspicuous

consumption, power and control. How do these large gender-related differences on the SV arise? How are they related to the effect sizes and percent-male statistics found by Benbow, Brody, Dauber, Lupkowski, Stanley, and others? Would it be in some sense desirable to try to change females' value orientations to make them more theoretical, economic, and political, but less social and aesthetic? Bibliographies included in the various Mental Measurements Yearbooks, originated by Oscar Buros long ago, could help one get started on this quest.

The effect sizes and upper-tail gender ratios reported in this symposium are meant to promote research of various kinds. We hope that some of you in the audience will respond to the implied challenge. Several of us in SMPY plan to examine the 82-test data more fully and probably explore factors related to some of the largest effect sizes.

Table 1: PHYSICS TESTS

(Listed in presumed decreasing order of difficulty)

	Effect Size	Percent Male
Graduate Record Examinations (GRE)	.56	85
Medical College Admission Test (MCAT)	.51	61
Advanced Placement Program Level C, Part 2, Electricity & Magnetism	.29	86
Advanced Placement Program Level C, Part 1, Mechanics	.37	85
Advanced Placement Progam Level B	.41	77
College Board High School Test	.59	81
Median	.46	83

Table 2: CHEMISTRY TESTS

(Listed in presumed decreasing order of difficulty)

	Effect Size	Percent Male
Graduate Record Examinations (GRE)	.57	63
Medical College Admission Test (MCAT)	.37	61
Advanced Placement Program (AP)	.33	70
College Board High School	.39	64
Median	.38	63.5

Table 3: HISTORY AND POLITICAL SCIENCE TESTS

(Listed in presumed decreasing order of difficulty)

	Effect Size	Percent Male
Graduate Record Examinations (GRE), Political Science	.76	64
Graduate Record Examinations (GRE), History	.57	64
Advanced Placement Program (AP), European History	.26	57
Advanced Placement Program (AP), American History	.24	53
College Board High School, European History	.58	65
College Board High School, American History	.40	58
Median	.485	61